Hypermedia = Hypercommunication

Hypermedia '90
December 5, 1990
Houston, Texas

Mark R. Laff

Interactive Media Project
IBM T.J. Watson Research Center, Hawthorne
P.O. Box 704
Yorktown Heights, New York 10598
BitNet: MRL at YKTVMH
CSNet: mrl @ ibm.com

Abstract

New hardware and software technology have given application designers the freedom to use new realism in human computer interaction. High-quality images, motion video, stereo sound and music, speech, touch, gesture provide richer data channels between the person and the machine. Ultimately, this will lead to richer communication between people with the computer as an intermediary. The whole point of hyper-books, hyper-newspapers, virtual worlds, is to transfer the concepts and relationships, the "data structure", from the mind of the creator to that of the user. In this talk we will discuss some of the characteristics of this rich information channel followed by some examples of our work, including an interactive hypermedia biography of IBM Fellow John Cocke entitled "John Cocke: A Retrospective by Friends".

Introduction

For the author, one of the most "visual" pieces of classical music is Tchaikovsky's "Peter and the Wolf". The musical themes representing each of the characters add vivid imagery to the story. Readers of this article will each have their own favorite example of music which evokes powerful associations. Some of these links are cultural and shared, while others are private and unique to an individual's own experience.

Moving to another medium, the adage "a picture is worth 1000 words" can be modernized to "a moving video picture is worth 30,000 words per second".[1] While this is simply a metaphor, it is certainly true that features of a person's movement can communicate a portion of the message. The phrase "yay high", for example, must be accompanied by the appropriate hand position to indicate just how high (or low) the speaker means. The dynamic aspects of body language augment the spoken dialog yielding a much richer information channel. Similarly, in the video and film media the motion of the camera is often used in a stylized fashion to convey a particular abstract concept to the viewer. A zoom-out may indicate an ending, while a zoom in may indicate a beginning.

These observations on communication are not new. What is new is the ability for computer software designers to utilize the power of these media in both existing and new computer applications. Digital signal-processing hardware has brought high-quality stereo audio, natural image, and motion video to the desk-top workstation. The challenge before us now is how to combine the power of these "natural" I/O media with the interactivity of the computer to yield more effective computer applications.

Not Just For Bank Balances, Anymore

Looking at the historical development and fundamental changes in computing, we find that the original uses for computer were numeric in nature. The computer kept track of a bank balance or solved a numerical integration. Hollerith strings stored character text, but were not the subject of computation. Input and output were typically boxes of punch cards and deep stacks of printouts, both containing rows and columns of numbers.

The next evolutionary step was to symbolic processing. While usually associated with Artificial Intelligence, we include as symbolic computation that which employs

1 30 frames (or images) per second is the standard television display rate in the U.S. Each frame actually consists of two different fields, each of which present alternating lines of an interlaced image, so at 60 fields per second we could say "60,000 words per second".

pointers. Thus, modern data structures and data bases are symbolic in nature, not just the symbolic programming languages (LISP, Prolog, etc.). The human-computer interface came to include names (symbols) and relationships. Powerful operations on the symbolic structures, such as search, inference, extrapolation, were the trade-offs for the extra space and time required to store and process the symbolic information.

This brings us to the present day, in which the new orientation is simply storage (recording) and retrieval (playback) of noncoded information. Although understood in some sense at a very low level, the natural images and sounds are simply translated between the analog and digital domains. There is no understanding of the information contained. Indeed, the objects of which we are speaking are so rich in content that humans rarely agree on the meaning of, for example, a picture of a wind-blown wheat field. The computer will happily "capture" that image and reproduce it when bidden.

It is in this context that we develop the notion of computer communication enhanced by this interactive multimedia technology. The point is that it is an "author" (nearly always a human, today) who defines the ordering and synchronization of this "playback" in order to convey a message. The "answer", formerly a number then followed by a list, is now a multi-sensory "experience". The richness of information content of the media allows "transmission" of the answer from the author to the user.[2]

An Example, Audio Annotation for Electronic Mail

Electronic mail is a ready example. Widely available from many sources, electronic mail is extremely effective for rapid communication between people, including both acquaintances and people who have never met. Arpanet mail and bulletin boards, for example, link people around the world in all manner of organizations (Universities, Industry, etc). A fundamental limitation of this communication[3] is the text-based nature of the information. While a usable least-common-denominator, the constraints have lead people to invent extensions mechanisms for augmenting the text. Out-of-band information, e.g. the instructions "----- cut here -----" within a message for separating the attachment from the base message is one example. A more apropos example is the invention of the character-based icons for conveying emotional tone. Character-grams such as :-) and :-(convey happiness (or humor) and sadness, respectively.[4] Even with these annotations, it has been the author's experience, as others have found, that written messages are all too easy to misinterpret. So much of the usual information which people depend on in conversation is missing (pitch, timbre, inflection, timing). It is a wonder that the message gets through at all.

Contrast the problems of text-based mail with the power of multimedia mail. One simple example of the possibilities, utilizing text, graphics, animation, and audio is the commercially available product Freestyle from Wang. Functionally, this product allows the user to capture a textual screen from any running application and add hand-drawn text and graphics along with audio annotation (usually speech). The recipient of the electronic mail message may then play back the annotation in "real-time". One hears the author's voice synchronized with the "ghost writing" exactly as they were recorded. Similar audio-annotated communication is available in the Macintosh and NeXT environments.

It did not take the author long to be convinced of the power and increased effectiveness of communication in this fashion. All the elements of normal telephone conversation are present,[5] plus the added dimension of the real-time handwriting playback. The visual clues, the body language, are all that are missing, and digital video compression hardware will pave the way for that, too.

Hypermedia Design

Given the potential capability of this communication medium, the question remains: "How do we use it effectively?". With a set of basic output elements (still and moving images, stored audio, text) and input devices (mice, joysticks, keyboards, touch screens, cameras, microphones) the number of possible combinations presents many opportunities for poor design. As Desktop Publishing made possible documents which use every font on every page, rendering it unreadable, there is potential for many unwatchable and unusable interactive hypermedia applications. We are reminded of the typical "home movie" - jerky images and awkward action

2 We feel that all of the current terms, "user", "viewer", etc., are inadequate for describing the person at the receiving end of the interactive multimedia experience. We have settled on "user" as a poor compromise.

3 Although it is hard to describe such a powerful medium as "limited".

4 To see this, tilt your head to the left to see the eyes : , nose - , and mouth) or (.

5 Except the interactivity, making this more like answering-machine-mediated communication.

("Everybody, wave at the camera!"). With the availability of consumer-grade video editing equipment, however, we find increasing sophistication in the story telling ability of home video. People are learning, with broadcast TV as a model, how to design in this new medium in order to "get the message across".

While this last example speaks to creating home videos, we feel that there are already-established schools of design for many media: video, film, graphics, drama. The introductory course for any of these domains, e.g. Graphic Design 101, teaches the basic principles for creating good designs.[6] With the addition of the interactive capability of the computer for hyper-link branching, we feel that a new set of design rules will arise from a synthesis of the rules from the various media which make it up. The result will be a course entitled "Hypermedia 101", and will address the issues arising from combination and coordination of the various output and input channels available, as well as issues of branching, including the underlying logical structure of the information being presented.

A key aspect to be treated in these guidelines is the development of involvement in the user - how to "draw" the user into the application. Such techniques already exist in various forms for the current media, for example "characterization" from film, stage, and literature. The creation of a persona with whom the user can identify and grows to care about is a powerful way to bring about involvement. Exploiting the "conversation" (interaction) as part of the user interface will be the challenge of design in this new medium.

Hyperchannel Communication

If an anthropologist from Mars were to land on Earth what theories could they derive from studying a current workstation or personal computer?[7] Imagine that all pictures of what we humans look like were mysteriously destroyed. What sort of creature would be re-constructed? It would have a very weak spine, requiring it to sit, constantly; monocular vision with very limited sensitivity to color; three hands (two for the keyboard, noting the symmetry of the left and right shift keys, plus one hand for the mouse) with very limited range of motion; and very poor hearing. We should ask why this reflection of ourselves in our technology is so far off base!

In this regard, perhaps hypermedia systems should instead be termed "hypersensory", as this is in fact one basis for the power and potential effectiveness of the technology. The justification for the increased cost of hypermedia in terms of storage and processing power is the more appropriate match-up in capabilities between the computer and the human. Observing what we, today, think of as a modern hypermedia application (with rich graphics, moving pictures, high-quality sound), our alien anthropologist would get a much more accurate picture of our sensory capabilities. If the application were also of the "virtual reality" genre utilizing stereoscopic "eye phones", stereo ear phones, a data glove, speech recognition/generation then the human portrayal would be much more accurate. This is not surprising, as virtual reality researchers[8] have often spoken of the explicit design goal of fully utilizing the human sensory capability in the user interface.

Experimenters at Xerox EuroPARC have been exploring audio output as part of an "Alternate Reality" environment. In the ARKola[9] experiment test subjects work jointly at different computers to run a simulated beverage-bottling plant. The graphical representation of the plant is manipulated so that the entire plant would not fit on the screen at one time. In the natural model which evolved for collaboration between the two subjects each focused their view on one-half of the plant and communicated between themselves to establish coordinated actions. In one test group of subject pairings, sound effects provided feedback on the state of operation of the plant. In the other test group, the application was silent. The finding was that using the sound of the operation of the non-visible portion of the plant did improve problem solving ability.

Each user receives cues on the actions (and their effects) of the partner which were directly coordinated with the operation of their portion of the plant. The extra audio information enhanced each user's internal model of the domain and improved the problem solving ability of the team. This example has obvious implications for cooperative work environments such as showing how the limitations of screen real-estate of the visual medium may be attacked using a multimedia approach.

As well as using the different I/O channels of a hypermedia system in parallel to convey different aspects of a single message, we may also use the various media for several different messages simultaneously. Without hypermedia, notification of asynchronous events, such as the arrival of new electronic mail, may be announced by popping up a window on the display

6 Of course sophisticated designers will break these rules and still be effective, but this is based on training, experience, and talent.

7 Thanks to Bill Buxton for this allegory.

8 Such as Jarad Lanier of VPL.

9 Demonstrated by Bill Buxton and Bill Gaver as part of the tutorial "Non-Speech Audio", Chi '90, Seattle, Washington.

screen. Given audio output, especially speech generation, the visual environment may be left undisturbed, using the aural channels to convey the notification. The choice of a particular channel gives the application designer new freedom to tailor the user interface to the semantic content of the data. While human capability for processing multiple distinct is limited, and it remains to be seen how fully this parallel communication may be exploited, simply removing the steps (keystrokes, mouse clicks, etc.) to close the notification window will be an improvement. Another very simple and compelling example[10] is the ability for the computer to choose voice communication over visual for the situation when the user is across the office not attending to the computer screen at all.

Another Example, John Cocke: A Retrospective by Friends

In our recent work, we[11] have been focusing on the effective combination of motion video (TV) with interactive branching. We undertook the John Cocke project as way of learning-by-doing what some of the parameters and boundaries are in design for this area of hypermedia. We are incorporating our discoveries pertaining to both application and tool design in our ongoing research.

Commissioned for a symposium honoring IBM Fellow John Cocke's 35th anniversary with IBM, we developed an interactive laser disc application depicting the man and his career. He has had a very rich history with IBM, has worked on many key advanced projects, and is recognized as originating many of the fundamental ideas behind compiler optimization, high performance computer design, the RISC architecture concept, among many others. He also has a unique personality and is warmly regarded (loved, actually) by all those that come to know him.

From the beginning, we felt that the interactivity of the computer combined with the power of video would help us with the difficult task of capturing the diversity and complexity of John Cocke. The history was told through video-taped interviews of 14 colleagues of John's, as well as John himself. Much in the style of a film or video documentary, we extracted the "choice" portions of the 22 hours of interview footage and boiled it down to a 1 hour laser disc.[12] Unlike a video documentary, however, the selected bits (termed "sound bites") were not woven into a single, linear piece. Instead, a hypermedia structure was designed to organize the video from several different perspectives:

Who by each interview, grouped by topics "The Man", "His Work", "Impact", "Style", "Stories".
What by major system project
Where places from John's history
When a time line by year
How how John does what he does, his personal characteristics
Why his importance to IBM, including key contributions, significant awards (e.g. the ACM Turing award)

Another top-level view was essentially a random organization. Called **Quiz**, this set of 35 trivia questions about John Cocke served both to give a general feel for the data (video clips) and also appealed to the entertainment aspect of the banquet and symposium. The multiple-choice questions in the quiz were formed by selecting interesting answers from the available material and then choosing the question to fit. An incorrect answer yielded a video of one of the subjects on the laser disc saying, "I'm afraid that's incorrect", or "You must mean X" where X is the correct answer. These positive, negative, and hint feedback pieces were taped long before the questions were designed. As with the answers, the available feedback shaped the design of the questions to some degree. Finally, a small section gave further explanation of the different perspectives (**Help**, for when our interactive design failed to be intuitively obvious) and described the underlying technology.

Presented using a touch-screen, laser disc, and video windowing adapter,[13] the user interface was styled as a tree of multiple choice menus with graphics and video stills combined. The leaf nodes of the tree consisted of video "sound bites". Early user testing suggested that it was important not to build the tree too deep, requiring a long sequence of menu choices to get to the "reward" video segment. We flattened the tree accordingly, taking at most three choices to reach a leaf. Testing also pointed out the motivational states for different users. Some users took more active control, navigating easily through the menus. Others preferred the information be presented to them by the system, with the user taking a more passive role.[14] This prompted us to expand the "attract mode" (as in video games) portion of the application, intended to play

10 Described by Nicholas Negreponti.
11 The Interactive Media Project, IBM Thomas J. Watson Research Center.
12 This selection and editing process is the subject of a set of papers currently in preparation.
13 IBM's M-Motion Video Adapter.
14 Users cited unfamiliarity with the information, and therefore no good basis for making navigation choices, as one reason for taking the passive role.

sequentially through the material when no one is using the kiosk. Originally a brief selection, the attract mode grew to cover much more of the material taking approximately 1/4 hour without repetition. The user may take control at any time simply by pressing the touch screen.

The result is an attractive, interactive presentation which gives a well-rounded view of the man, John Cocke. The user receives a view of his career with IBM, his impact on not just IBM but the entire field of Computer Science, and also very warm, personal accounts of his unique style. The video interviews with his colleagues and friends gave a richness in variation and historical feel to the account, a "Retrospective". The number of people we interviewed, many of whom used surprisingly similar terms to describe John, helped give weight to the individual comments. And, finally, the use of video clips gives the feel that the people on the laser disc are speaking directly to the user.

Conclusions

We in hypermedia research are exploring the use of the recently available digital hardware which brings rich analog media to the desktop. The challenge lies in designing for this new communication medium which borrows from film, TV, literature, and adds interactivity or branching. The potential power of the richness and realism in the user interface will provide an enhanced communication channel between designer and user; user and peer user; computational model and user. The input ability of the channel will allow consideration of the user's state of mind, such as a history of the recent screen touches, a joystick input device for continuous indication of the user's interest level, or visual sensing of the user's body language. We will learn, over time, how to design systems which better match the sensory capabilities, vision, motion, sound, touch[15] of the people who use them.

[15] Some day perhaps taste and smell, too.